DOCUMENT RESUME

ED 242 658 SP 023 706

AUTHOR Williams, Robert T.
TITLE Program Evaluation through Informal Comparisons.
PUB DATE 3 Feb 84
NOTE 18p.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Admission Criteria; Decision Making; *Educational Change; Evaluation Methods; Higher Education; Policy Formation; Preservice Teacher Education; *Program Evaluation; *Program Improvement; Schools of Education; *Self Evaluation (Groups); Teacher Education Programs; *Teacher Supply and Demand

ABSTRACT

Recommendations are made for improving teacher education starting with internal reform on the part of individual schools of education. It is pointed out that institutions, acting independently and comparing their programs in a realistic way with similar programs in other schools, can identify weaknesses and are in a position to rectify them. An example is given of the problem of teacher supply and demand, with shortages in some teaching fields and surpluses in others. It is noted that this is an indication that many institutions are following a pattern that contributes to this imbalance. An institution, it is pointed out, has an option to reduce its production of graduates in the surplus field and increase production in the shortage field. It is suggested that, by cutting out weaker students in fields of surplus, the following benefits would be seen: (1) an increase in the quality of the remaining graduates; (2) a more favorable student-teacher ratio; (3) an opportunity to redirect or reduce teacher education expenditures; (4) an increase in morale and self-respect in teacher education students and faculty; (5) a reduction in the number of graduates who can't get a job in their field of certification; (6) a reduction in the number of teachers available for out-of-field assignments; (7) a reduction in the number of teachers who perform poorly; and (8) an increase in the image of teacher education. (JD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED242658

February 3, 1984

# PROGRAM EVALUATION THROUGH INFORMAL COMPARISONS

Robert T. Williams
N. C. State University

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

✓ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Robert T. Williams

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SP023706

February 3, 1984

# PROGRAM EVALUATION THROUGH INFORMAL COMPARISONS

Robert T. Williams
N. C. State University

The critics of teacher education are not using the same standards for program evaluation that teacher educators have been using. But the standards used by our critics are acceptable and meaningful to the media, to decision-makers in political positions, and to the general public. Because we are not autonomous, and our graduates are employed in the public sector, the critics' standards must be accepted by teacher educators, and used in our program evaluation, if we expect to regain public support and respect. An article in the October 19, 1983 Education Week is an example of this. It was reported from Texas that the new state-mandated basic skills admissions test would seriously cut teacher education enrollment at U. T.-EL Paso. My reaction is, "Thank God!" Merely the fact that the basic skills of teacher education students is a nationwide issue of our critics is an indication of our failure to have appropriate evaluation processes.

It is clear that teacher education programs must change -- not just in ways that we would like to see them change, and at our speed -- but in ways that satisfy our critics. For those who can't change in this way, they had better get out of the business.

At last year's AACTE meeting in Detroit, I attended a session on program evaluation that left me livid. There was a lot of talk about evaluation models, changes in NCATE standards, and jargon about process, context, internal, summative and other kinds of evaluation. But I came away with nothing that I could use or build on to evaluate programs at my institution. I also came away with a sense of despair. It seemed to me that we were

SP023706

fiddling around while teacher education was under attack and being burnt to the ground.

I don't have much faith in formal models of program evaluation. Maybe it's simply because I don't have much understanding of them. But it seems to me that there's seldom an obvious connection between a model and the detailed activities needed to use the model.

I don't have much faith in the self-study, visiting team process of evaluation, either. In our state, at least until last year, this process was characterized by the good-ole-boy approach. Everyone was approved. Most of the right questions aren't even asked in this process.

It's clear that traditional program evaluation processes have seldom resulted in identifying weak programs, and in identifying weak aspects of satisfactory programs, and in correcting these weaknesses. I have stopped being a member of visiting teams, but I can remember being told by one chairman that our charge was only to verify what the institution said in its self-study. The ultimate to this is for the institution to say that its programs are very weak and for the visiting team to say, "Yes, we agree." But that doesn't guarantee that improvements will follow.

Our governmental and institutional decision-making processes seldom follow grand-plan, long-range plan methods. Our decision-making processes are described by political scientists as disjointed incrementalism. We make a little change here and a little change there, one change this year and perhaps one the next.

The wise teacher educator does not wait for a visiting team or for other outsiders to identify weaknesses. Each of us needs to become sensitive to where we need to make changes, and to become assertive in making them. Teacher education will gain the respect of its critics only if hundreds of us are simultaneously making significant changes where we can, when we can. The accumulation of these changes may save us before all control and decision-making is taken from us. We know from experience that if we don't make the improvements that we have the authority and responsibility to make, then others will step in and mind our business for us.

In the December, 1983 AACTE Briefs, President Flowers called for this annual meeting to "...launch an era of teacher educator involvement, action, and assertiveness in educational reform." To me the key word there is assertiveness. Other articles in the same Briefs cited efforts of the Texas ACTE and Minnesota's MACTE to take this leadership. Our North Carolina affiliate is making some moves in the same direction. I hope that these initiatives are not too late.

In a short, but pointedly satirical paper, Wolf caricatured educators' evaluation efforts by sorting them into five categories (see attachment). This is his description of the colloquial method:

> Social psychological research has demonstrated that decisions arrived at by a group will achieve greater acceptance than decisions arrived at by an individual. This finding is the basis of the colloquial method. In applying this method, one need merely assemble a group of people who have been associated with a particular program to discuss its effectiveness. After a brief discussion, the group will usually conclude the program has been indeed successful. This conclusion can then be transmitted to funding agencies and other school personnel. It is unlikely that such evaluations will be challenged since they have been arrived at by a group.

Again, this reminds me of the visiting team process.

I like a phrase that David Imig used at the 1983 AACTE Summer Institute: The quantification of quality. Quantitative data frequently present opportunities for qualitative judgments. But we often neglect to use data that are readily available. I sometimes wonder whether that is by design, because to find and admit to shortcomings implies an obligation to change. But more about that later.

In this paper, I am defining the word evaluation as a process for determining the value of something; and a process for making judgments. The judgments often come down to simplistic statements, such as: This is good; this is adequate; this is bad;this must be improved; and,this must be stopped.

These judgments are most often made because we have compared things. They are relative judgments. Sometimes the things that are compared are numerical, e.g., rankings of grade point averages, or averages of test scores, or numbers of teachers produced. Sometimes the comparisons are not numerical, e.g., impressions gained through interviews, observations of student teachers, and feedback comments regarding the performance of graduates, or of their perceptions of your program's strengths and weaknesses.

With the critics barking ~~out~~ at our heels, it is no longer sufficient to say that the purpose of evaluation is not to prove, but to improve. In my world, the purpose of evaluative judgments is to make decisions regarding the need for program change. A change may be immediate, short-range, or long-range. It may have to wait for new dollars, or may be made without new money. The change may have to wait for one or more people to resign or

retire. The effective manager has often decided what changes need to be made before the opportunity to make a change presents itself.

My concept of change is broad in that it encompasses many administrative levels, from the national level down to the classroom. We know that there is substantial variation in the quality of teacher education. At the national level, there is some variation between the states. Within each state, there is some variation between institutions. Within institutions, there is some variation between programs, and within programs there is qualitative variation between students. Let's use the qualitative indicator of SAT scores to illustrate. Now assume that the mean SAT score of all college students is 1000. The distribution may or may not be a normal curve; it doesn't matter. Within that set of all college students is a subset of all teacher education students. That subset also has a mean SAT. Let's assume that it's 930. Again the shape of this distribution is not so important. An appropriate question is, is this discrepancy between the two means ok? What I hear and read tells me NO.

Within the subset of all teacher education students is the subset of North Carolina's teacher education students. And Arkansas'. And South Dakota's, etc. There is a mean SAT score for the teacher education students in each state. Compare your state mean to the national mean and ask, "Is this discrepancy ok?" The next subset contains each teacher education institution in your state, each of which has a mean SAT. Within institutions, we use the mean SAT for each program. And finally, within each program we use the SAT score for each student.

At each level where we are not satisfied with the discrepancy, we must eliminate the lower end of the tail, by denying admission or by quantum improvement in the performance of those students in the lower group. We can also raise the mean by admitting more students at the upper end of the tail, but we have less opportunity for this. Most of us do have the opportunity to affect the quality of teacher education at the student, program, and institutional levels. Each institution must set its own objectives. A few of us will have the opportunity to affect quality at the state and national levels. Our personal self-respect, our institution's reputation, and perhaps our jobs ride on the actions we take -- or don't take.

At North Carolina State University, data are published each year on the SAT scores of the freshman class, by schools. Here are our average SAT scores for the past ten years: fairly steady, usually higher than our assumed mean score for teacher education students nationally, but lower than our assumed mean scores for all college students nationally. The School of Education has consistently ranked below the mean for our institution. Each blue column represents one of the other schools on campus. In 1973, we were on the bottom, and 110 points under the campus mean. In 1983, we had passed one school, and were close behind two others but still 80 points under the campus mean. Is this discrepancy ok? Our objective is to be at the campus mean, or above it. Again, we can move towards this goal by raising the floor and/or by attracting more students with higher SAT scores. For the past several years, we have worked at raising the floor. Beginning last year, we began to court more students with higher SAT scores. We are now gathering data to enable us to make SAT comparisons of our graduates, as well as of our new freshmen. There

is reason to think that the mean SAT score of our seniors is higher than that of the same class when it entered, because we receive a substantial number of transfer students, and because some of our weakest freshmen never graduate. If we find out that those transferring in are holding down our mean score, we may decide to raise our transfer requirements.

Another level where we can (and do) make evaluations based on informal comparisons is between programs within the School of Education. Each of us has a weakest program in our school and a strongest program in our school. We base these evaluations on both quantitative and non-quantitative inputs. We then need to ask the questions: Is the weakest program too weak? If yes, what can we do to strengthen it? In our case, the weakest program ten years ago was one that had become a haven for scholarship athletes -- more specifically, athletes who were poor students. In the fall of 1976, the School of Education enrolled 96 new freshmen. 23 of them were in this curriculum. 21 of the 23 were on an athletic scholarship. 21 of the 23 were admitted as exceptions to the University's admission policy. Only three of the 23 graduated. The only one who is still attending is scheduled to graduate in May. Notice on the transparency the enrollment in this curriculum for the past fifteen years. It has decreased from a high of 99 in 1976 to a current low of 20. Many people would judge this sharp enrollment decline to be bad. I judge it to be both good and bad, but mostly good. After the fall of 1976, we slammed the door on the athletes who were poor students. It took five years for this action to have its effect, while those students worked their way out of the institution. The good part is that this type of student is no longer a substantial component in the School of Education.

The bad part is that we have not yet attracted better students into the program to meet the need for teachers in this field. We often find that quality is inversely proportional to quantity.

After satisfying ourselves that this program was under control, we turned to a program that seemed to be the next weakest. The criteria we used were student NTE scores and the fact that a noticeable number of students were meeting their chemistry requirement at community colleges during the summer, instead of at N. C. State. Our treatments were to push for a higher GPA admission requirement to this curriculum and to insist that the students at least try chemistry at our institution before enrolling for it elsewhere.

In both of these cases, we are using the strategy of strengthening the weakest link in our school. We have applied this not only to student quality, but also to faculty performance. Although the results on the faculty side come much slower, that is no reason not to apply pressure to the weak links.

Evaluation does not always have to use data that are qualitative. There are some possibilities for informal comparisons with respect to teacher supply and demand. We know that at the national level there have been persistent shortages in some teaching fields and persistent surpluses in other teaching fields. We have regional and state-level data as well. This can only mean that many institutions have the same pattern. Does your institution contribute to this national imbalance? If yes, has your institution acted in the last three-to-five years to reduce the imbalance? An institution which contributes to the imbalance has several choices:

1. It can fail to acknowledge the situation.
2. It can acknowledge the situation but fail to act.

3. It can act to reduce its production in the surplus field.
4. It can act to increase its production in the shortage field.
5. It can do both #3 and #4.

Clearly, #1 and #2 are the easiest courses to follow and most of us have done this, even though they have their disadvantages. Alternatives three through five also have some disadvantages -- and maybe some pain. But the advantages they offer include respect for attacking a problem. To me, this is an example of what President Flowers meant by assertiveness in education reform.

In Educating a Profession, the Bicentennial Commission recommended, "That the AACTE begin to take strong stands on teacher education and other educational issues, even when the stands are contrary to the interests or desires of some member institutions." Supply-and-demand problems are one such issue. But even in the absence of such leadership at the national level, we can address such issues at the state level through AACTE-affiliates, the state board of education, or other agencies. And even in the absence of leadership at the state level, 200 institutions acting independently can make a tremendous correction of the imbalance.

If all we did was to cut back in fields of surplus, but cutting out our weaker students, we should see the following benefits:

1. An increase in the quality of our remaining graduates;
2. An opportunity to have more favorable student-teacher ratios in teacher education;
3. An opportunity to redirect or reduce teacher education expenditures;
4. An increase in morale and self-respect in the students and faculty remaining in teacher education;
5. A reduction in the number of frustrated teacher education graduates who can't get a job in their field of certification;
6. A reduction in the number of teachers available for out-of-field assignments;
7. A reduction in the number of teachers who perform poorly, and who thereby are poor and unfair advertisements for teacher education; and
8. An increase in the image of teacher education.

Most of us already know if our institutions are contributing to the supply-demand problem, or are part of the solution. The data for comparisons are available.

We indirectly participate in lowering the quality of the teaching force when we continue to graduate far more new teachers of average or marginal quality than are needed in certain fields. In many states and school districts, such persons are hired to fill positions out-of-field. If we restricted our enrollment in fields of surplus (by putting a lid on the number of graduates per year, and by raising standards), we would cut off this oversupply which is used by school systems to cover up shortages. We would help the public see a true condition of practice, and thereby enhance change.

This is a price that some of our member institutions must pay if we as an Association are to share in the accountability for the quality of the public school teaching force. The public does not know when out-of-field assignments are rampant in their local schools. The teachers aren't going to tell them; they want the job. The local administrators aren't going to tell them; they want to keep things peaceful. So when Johnny's out-of-field math teacher is ineffective because he's not competent in mathematics, the parent questions the effectiveness of the teacher's training. Why should AACTE-member institutions continue to take a bum rap for the unethical practice of out-of-field teaching? No wonder we don't get respect. This is an example of what Howsam, et., al., referred to as becoming more consumer-oriented and less practitioner-protective, in Educating a Profession (AACTE, 1976).

One other place to direct our evaluation efforts is to the relationship between our purpose, responsibilities, and capabilities. Educators including teacher educators, have made the mistake of trying to be all things to all people, when we didn't have the resources to accomplish many of the responsibilities that we bit off. I hear this when I hear some of the AACTE leaders talk about changing "the conditions of practice." I think they are talking about working conditions and practices in elementary and secondary schools. But we are not accountable for such working conditions and practices in an institutional aggregate sense, and we are not funded to attack and change these things. So let's back off from this commitment. However, we are accountable in the sense that each of our individual graduates had better be a contribution to improving current classroom practices.

When public schools are not able to hire those of our graduates they would like to, this tells us that their positions are not competitive. We know that the two major reasons why they are not competitive are salary and working conditions. Neither AACTE nor individual IHEs can "cure" these two problems. Only legislatures, boards of education, and state and local school personnel can cure these problems. We can supply them with relevant and useful information, but we cannot make them act, and we need not feel accountable nor responsible for these conditions.

The examples I have chosen to discuss are only a few of those available for evaluating through informal comparisons. Here is a list of some others that come to mind:

1. Pass rates on output measures, such as the National Teacher Examinations. I have brought a handout that shows such a comparison for teacher education institutions in Louisiana;

2. Grade point averages. Two years ago, the School of Education at the University of Wisconsin-Milwaukee compared grades earned by its students with those earned by other students on campus who took the same courses, and showed that School of Education students were achieving at a rate higher than the campus average.
3. Faculty loads;
4. Faculty publications;
5. The teacher education funding formula; and
6. The control of teacher education on campus.

Each of these items is fodder for comparison, and for evaluative judgments based on the comparisons.

Since there are good reasons to constantly use evaluation to make improvement decisions, why don't we see more change more quickly? There are many answers. One answer is that the greatest number of teacher educators --the full-time teaching faculty--have a limited view of the big issues and have limited opportunity to make necessary changes. When I said earlier that hundreds of institutions acting independently could make a big change, that implies that several faculty members at each of these institutions would be involved in agreeing on the problem to be addressed and on working to implement the change. But the work ethic at many institutions rewards working independently, to further develop one's area of expertise.

The teaching faculty have some other very good reasons for not making more changes more quickly: most of them are not in a position to influence resources shifts; many of them are not trained in policy analysis; some of them probably don't even understand the problem; their self-interest may be contrary to the greater interest that I have been addressing. In addition, there are several aspects of their faculty-student interrelationships

that deter qualitative changes. I have listed over 20 of these factors on a handout. This is another example of the concern that we need to become more consumer-oriented and less practitioner-protective.

If we cannot expect the teaching faculty to provide the leadership for change that is needed, then it falls to the administrators in teacher education--department heads and deans--to provide that leadership. Again there are several reasons why we don't see more change more quickly: lack of training in policy analysis; inability or unwillingness to admit the problem; laziness; low personal standards; a narrow or short-term view of what one's self-interest is, a lack of knowing how to go about solving the problem, etc.

For many of us, the answer is not so much knowing what to do as in having the courage to do it. Where this is the real problem, let's start with the possible. That means that those who wish to be responsible must be honest about where the problems lie. As Pogo said once, "We have met the enemy, and he is us." We can create some opportunities to help our weak institutions get stronger (such as Operation Share, in North Carolina), while at the same time putting them on notice that without improvement they will not continue to receive our support. The same goes for weak departments and programs in otherwise acceptable institutions. The same goes for weak faculty and students in otherwise acceptable departments. If you have gangrene, you treat it successfully or cut it off; you don't just ignore it while it grows worse and eventually kills you.

The time is ripe for reform. University administrators are in a mood to strengthen teacher education. The public and the political leaders will support improved standards. This is the opportunity of a career; let's capitalize on it.

We expect a widespread teacher shortage within ten years. If we get our house in order now, and hold the quality line when the shortage hits, our respect will be restored. Then we will attract more of the type of people who should be in teacher education. People with ambition and high standards want to be associated with a winner, not with a loser.

## THE 5C EVALUATION MODEL

| | |
|---|---|
| 1. The Cosmetic Method | Does it look good? |
| 2. The Cardiac Method | Do you believe in your heart that it is good? |
| 3. The Colloquial Method | If a group of associates says that it is good, then it must be good. Who will argue with a committee? |
| 4. The Curricula Method | There is an inversely proportional relationship between the effectiveness of a program and the disturbance of schedules, pupil placements, and teacher habits. |
| 5. The Computational Method | Get two clerks to process the same data. The extent to which they get the same answers is a measure of correlation. Who can argue with a high correlation? |

Wolf, Richard, "A Model for Curriculum Evaluation." Psychology in the Schools, Vol. 6, #2, April, 1969

NATIONAL TEACHER EXAMINATIONS PERFORMANCE OF
LOUISIANA INSTITUTIONS OF HIGHER EDUCATION

| Institution | 1978-79 | | 1979-80 | | 1980-81 | |
|---|---|---|---|---|---|---|
| | PASSED | FAILED | PASSED | FAILED | PASSED | FAILED |
| 1. | 254 | 57 | 203 | 41 | 189 | 35 |
| 2. | 161 | 69 | 110 | 55 | 118 | 43 |
| 3. | 123 | 88 | 102 | 83 | 93 | 50 |
| 4. | 162 | 30 | 157 | 16 | 127 | 12 |
| 5. | 104 | 77 | 83 | 33 | 72 | 30 |
| 6. | 99 | 77 | 86 | 43 | 87 | 33 |
| 7. | 124 | 48 | 76 | 29 | 80 | 15 |
| 8. | 102 | 69 | 113 | 53 | 95 | 49 |
| 9. | 8 | 148 | 9 | 139 | 18 | 51 |
| 10. | 85 | 53 | 71 | 34 | 42 | 16 |
| 11. | 2 | 132 | 3 | 37 | 14 | 55 |
| 12. | 49 | 15 | 34 | 8 | 39 | 9 |
| 13. | 41 | 21 | 37 | 8 | 19 | 5 |
| 14. | 33 | 1 | 40 | 5 | 27 | 2 |
| 15. | 4 | 26 | 7 | 51 | 8 | 27 |
| 16. | 24 | 4 | 12 | 4 | 11 | 5 |
| 17. | 21 | 6 | 26 | 6 | 22 | 2 |
| 18. | 20 | 4 | 22 | 11 | 13 | 4 |
| 19. | 23 | 0 | 49 | 0 | 22 | 1 |
| | 1439 | 925 | 1246 | 656 | 1096 | 444 |